A Specification for a Highly Parallel Computer System

David Bailey, John Barton, Russell Carter,
Thomas Lasinski, Horst Simon[1]

Report RND-91-015, December 1991

NAS Systems Division
NASA Ames Research Center
Mail Stop 258-6
Moffett Field, CA 94035-1000

---

[1] NASA Ames Research Center, Moffett Field, CA 94035-1000

A Specification for a Highly Parallel Computer System
RND-91-015

David Bailey, John Barton, Russell Carter, Thomas Lasinski, Horst Simon
Numerical Aerodynamic Simulation
NASA Ames Research Center
Moffett Field, CA 94035, USA

ABSTRACT

A specification for a highly parallel computer system is presented. The specification is designed to facilitate acquisition of the highly parallel computer system through the use of a competitive procurement. The system specification is designed to cover all aspects of a production level supercomputer system, including floating point architecture and performance issues; main memory and mass storage capacity and performance; operating system and programming languages software support, and network hardware and software interface and performance requirements. The floating point computational rate is evaluated primarily on the basis of performance achieved on the NAS Parallel Benchmarks.

1 INTRODUCTION

The purpose of this specification is to define the requirements necessary to acquire the NAS HPCCP Testbed 1 (NHT-1) computer system, which will be installed in the Numerical Aerodynamic Simulation (NAS) system at NASA Ames Research Center, Moffett Field, California. It is mandatory that offerors meet all of the requirements defined in Sections 2 through 7 below. In the event an offeror cannot meet these requirements, the offeror must supply information describing what can be provided and projections of when the desired feature would be available.

1.1 Background

The NAS Program is a large scale effort to advance the state-of-the-art in computational aerodynamics. This facility is one of the premier supercomputer centers in the world and provides scientific computational services to a wide range of local and remote users representing NASA, other government agencies, academia and industry. The program has been tasked with providing research support for the High Performance Computing and Communications Program (HPCCP) leading to the development of computer systems capable of performing complex scientific computations at a sustained rate greater than current generation supercomputers. The long term plan of the NAS Program is to make available to its users an advanced, highly parallel computer system capable of achieving one trillion floating point operations on a sustained CFD calculation, by the year 2000. The installation of the NHT-1 system capable of a sustained computation rate of 3-5 Gigaflops will be an important milestone in attaining this objective.

1.2 Description of Environment

The NAS Processing System Network (NPSN) currently consists of a Cray Y-MP system, a Cray-2 system, a Connection Machine-2 system, an Intel iPSC/860 system, as well as numerous other auxiliary systems that provide for interactive usage, networking and mass storage. Through its long-haul communication system, scientists across the United States access the NPSN via high speed data links.

1.3 Definitions for Hardware/Software Specifications

In the following, the term "processor" is defined as a hardware unit capable of executing both floating point addition and floating point multiplication instructions. The "local memory" of a processor refers

to randomly accessible memory that can be accessed by that processor in less than one microsecond. The term "main memory" refers to the combined local memory of all processors. This includes any memory shared by all processors that can be accessed by each processor in less than one microsecond. The term "mass storage" refers to non-volatile randomly accessible storage media that can be accessed by at least one processor within forty milliseconds. A "processing node" is defined as a hardware unit consisting of one or more processors plus their local memory, which is logically a single unit on the network that connects the processors. The term "computational nodes" refers to those processing nodes primarily devoted to high-speed floating point computation. The term "service nodes" refers to those processing nodes primarily devoted to system operations, including compilation, linking and communication with external computers over a network.

The offeror's response shall specifically state which part of the proposed system consists of computational nodes and which part consists of service nodes. These two subsystems need not be mutually exclusive; system nodes may also be computational nodes, and vice versa. However, the requirements listed below for computational nodes must be met by the subsystem identified by the offeror as computational nodes, and the requirements listed below for service nodes must be met by the subsystem identified as service nodes.

2 NHT-1 SPECIFICATION

The offeror is requested to prepare proposals for Testbeds corresponding to two funding levels. The first funding level is approximately $5 million dollars and a system proposed at this level will be denoted a "Half-Size Testbed." The second funding level is approximately $10 million dollars and a system proposed at this level will be denoted a "Full-Size Testbed." In the following, specifications not explicitly targeted to either the Half-Size Testbed or the Full-Size Testbed implicitly apply to both, generically denoted the "NHT-1."

The contractor shall provide a testbed system with a sustained floating point computation rate on certain CFD-based benchmarks of 1.5 billion floating point operations per second for a Half-Size Testbed, or 3 billion floating point operations for a Full-Size Testbeds.

The Half-Size Testbed will have a minimum of 4 billion bytes of main memory and a minimum of 16 billion bytes of mass storage. The Full-Size Testbed will have a minimum of 8 billion bytes of main memory and a minimum of 32 billion bytes of mass storage. Details of the performance requirements are given in Appendix A.

2.1 General Hardware Specifications

2.1.1 The Half-Size Testbed must include at least 64 computational nodes. The Full-Size Testbed must include at least 128 computational nodes.

2.1.2 Each computational node must be capable of performing floating-point arithmetic on numbers in the 64 bit IEEE-754 "double" format, as specified in [8.16].

2.1.3 On the computational nodes, the results of addition, subtraction, and multiplication of 64 bit floating point numbers must be accurate to at least 52 bits, as specified in [8.16].

2.1.4 The hardware shall be capable of reporting all detected processor errors.

2.1.5 In the event of an underflow on a computational node, the node shall be capable of either flagging an error, or at the user's discretion, setting the result to zero and continuing without intervention by system software or causing a processor interrupt.

2.2 Main Memory Specifications

2.2.1 The size of main memory for the computational nodes of the Half-Size Testbed shall be at least 4 billion bytes. The size of main memory for the computational nodes of the Full-Size Testbed shall be at least 8 billion bytes.

2.2.2 The offeror shall provide the option to increase the size of the main memory for the Half-Size Testbed to at least 8 billion bytes. The offeror shall provide the option to increase the size of the main memory for the Full-Size Testbed to at least 16 billion bytes.

2.2.3 An individual computational node must be capable of correcting any single bit error in data fetched from or stored to its local memory and must be capable of detecting any double bit error. The computational processor network must be capable of correcting any single bit error in data communicated between nodes and must be capable of detecting any double bit error.

2.2.4 The memory system shall have the capability of reporting all detected main memory and interprocessor communication errors, both correctable and uncorrectable.

2.3 Mass Storage Specifications

2.3.1 The Half-Size Testbed shall have at least 16 billion bytes of mass storage. The Full-Size Testbed shall have at least 32 billion bytes of mass storage.

2.3.2 The mass storage system shall have the capability of reporting all detected mass storage errors, both correctable and uncorrectable.

2.3.3 There shall be a reliable backup mechanism for data stored in the mass storage system. Backup across an external network is not satisfactory. Transfer rate from the mass storage system to backup media must meet or exceed 8 megabytes per second.

2.4 Data Communication

The NHT-1 shall include a data communication capacity sufficient to support the simultaneous operation of the mass storage system described above, at least four (4) Ethernet network interfaces, and at least four (4) HiPPI/Ultra network interfaces.

2.4.1 The NHT-1 shall provide the capability of attaching at least four (4) HiPPI interfaces [8.15]. Each of these interfaces must also provide access to an Ultra network [8.8] (HiPPI/Ultra). These HiPPI/Ultra interfaces shall be capable of transferring at least 40 million bytes per second per pair of adapters in a sustained loopback mode over the transmission media. The source and destination files shall be resident on the mass storage system required in Section 2.3. This loopback test shall be performed between one pair of adapters. The details for testing this requirement are in Appendix A, Section A.5.3.

2.4.2 The NHT-1 shall provide the capability of attaching at least four (4) Ethernet interfaces. These Ethernet interfaces shall be capable of transferring at least 0.250 megabytes per second per pair of adapters in a sustained loopback mode over the transmission media. The source and destination files shall be resident on the service node file system. This shall be performed between two pairs of adapters. The details for testing this requirement are in Appendix A, Section A.5.3.

3 NHT-1 SOFTWARE SPECIFICATIONS

3.1 System Software Requirements

3.1.1 Source Code

At installation, the contractor shall provide, if requested, magnetic tapes containing all of the source codes for the current system software. Vendor will also allow access to on-line source on Government request at any time.

3.1.2 Operating System and Utilities

3.1.2.1 The operating system interface on the service nodes shall be in conformance with UNIX System V, Release 4 as defined in reference [8.1]. This operating system shall pass the System V Validation Suite (SVVS) published by AT&T Bell Laboratories, as specified in [8.1].

3.1.2.2 In addition to the standard UNIX System V system commands, facilities shall be available on the service nodes to allow a user to initiate, terminate, suspend, resume and adjust the priority of a computational job, and to lock a computational job into a specific node or set of nodes.

3.1.2.3 The operating system on the service nodes shall log and report all errors detected by the hardware, including single bit and double bit main memory errors on the computational nodes, interprocessor network communication errors, and disk errors.

3.1.2.4. On nodes that provide network access, a number of Berkeley UNIX 4.3 commands (as defined in reference [8.2]) shall also be included as follows:

3.1.2.4.1 The socket interface, as described in Sections 2, 3, and 4 of [8.2] is required. This includes the system calls *accept, bind, connect, gethostid, gethostname, getpeername, getsockname, getsockopt, recv, recvmsg, recvfrom, select, send, sendto, sendmsg, shutdown,* and *socket.* In addition, modifications to *read* and *write* to allow socket operations shall be implemented as described in Section 2 of [8.2].

3.1.2.4.2 The library routines described in Section 3n of reference [8.2] are required.

3.1.2.4.3 The "special files" described in Section 4 of [8.2.] that pertain to ethernet and Ultra support, as well as the interfaces described in the *arp, inet, ip, pty,* and *tcp* manual pages shall be supported.

3.1.2.4.4 The programs *hostid, hostname, optimize, ftp, ftpd, sendmail, telnet,* and *telnetd,* as described in Sections 1 and 8 of reference [8.2], shall be supported.

3.1.2.5 The socket interface shall support the Internet protocols IP, ICMP, TCP, FTP, SMTP, UDP, and TELNET, as described in reference [8.3] for all nodes that provide network access. In addition to this, all networking functionality set forth in references [8.9] and [8.10], shall be provided.

3.1.2.6 Subnetting: The system software on nodes that provide network access shall comply with RFC 950, as described in reference [8.6].

3.1.2.7 NFS: The RPC, XDR, and NFS client and server code from SUN Networked File System shall be provided on the service nodes by the vendor, as specified in references [8.11], [8.12], and [8.13].

3.1.2.8 X11: Either a client implementation of X-windows shall be provided or a client implementation of NeWS with X11 emulation shall be provided on the service nodes by the vendor, as specified in reference [8.14].

3.1.2.9. The operating system on the service nodes shall support the Network Queueing System (NQS) to allow submissions of jobs from other NPSN nodes as specified in reference [8.5].

3.1.2.10 The system shall allow a single job to utilize all computational nodes and at least 80 percent of main memory. This operating system configuration must be identical to that under which the Multiple Processor Performance Test is run.

3.1.2.11 The system shall have an efficient facility to allow at least twenty users to simultaneously run jobs on the computational nodes. It shall also allow the allocation of nodes between different job sizes to be changed without restarting the operating system.

3.1.2.12 The system shall have an efficient facility to prevent a user from accessing computational nodes assigned to another user or the operating system.

3.1.2.13 If the system allows more than one user to share a single computational node, it must include an efficient facility to prevent

one user from accessing the data of another user or the operating system.

3.1.2.14 The maximum file size allowed for a job on the computational nodes must be at least twice the size of main memory. This operating system configuration must be identical to that under which the Multiple Processor Performance Test is run.

3.2 Programming Software Support

3.2.1 Compilers for the Fortran-77 language, as defined in reference [8.4], shall be supplied on the service nodes. One shall compile programs to run on the service nodes, and one shall compile programs to run on the computational nodes. These compilers may be combined in one. The offeror shall certify that these compilers pass the Fortran-77 validation suite of the Federal Software Testing Service. Any preprocessor used to analyze Fortran source code is considered in this document to be part of the compiler.

3.2.2 Compilers for the C programming language, as defined in reference [8.7], shall be supplied on the service nodes. One shall compile programs to run on the service nodes, and one shall compile programs to run on the computational nodes. These compilers may be combined in one. The offeror shall certify that these compilers conform to the ANSI draft standard for the C programming language, as specified in the current draft standard in reference [8.7]. The C compilers shall include support for the entire C subroutine library, as described in the ANSI draft standard for the C programming language in reference [8.7]. Any preprocessor used to analyze C source code is considered in this document to be part of the compiler.

3.2.3 In addition to these standard features, the Fortran and C compilers for the computational nodes shall support parallel computation, if necessary using constructs or extensions. They must allow a single Fortran or C program to utilize all computational nodes and at least 80 percent of main memory. Additionally, high-speed asynchronous I/O between the computational nodes and the mass storage system must be supported.

3.2.4 The maximum file size supported by the Fortran and C compilers for a single job on the computational nodes shall be at least twice the size of main memory.

3.2.5 Codes written in assembly language or other languages for the computational nodes shall be able to access Fortran and C subroutines,

and vice versa. Fortran programs shall be able to call C subroutines, and vice versa.

3.2.6 The vendor shall supply a symbolic debugger, with a dbx-like interface, that will work on Fortran and C codes running on the computational nodes.

3.2.7 The same loader shall be used for all languages designed for programs running on the service nodes. The compilers for these languages shall all produce AT&T ELF object modules as defined in reference [8.19], and shall include all ELF debugging information as a compile time option.

3.3 NHT-1 Support Requirements

3.3.1 Handling of System-Related Problems

The contractor shall provide a clearly defined single point of contact for reporting system problems twenty-four (24) hours per day. A mechanism shall be provided for prompt handling of system-related problems reported by the Government. This mechanism shall provide for regular communications to the Government of both the status of these problems and the contractor's plan for dealing with them. The mechanism shall also provide for response escalation depending on the severity of the problems.

3.3.1.1 The contractor shall provide access to the Government of hardware monitor reports on the NHT-1 system.

3.3.2 Personnel Support

3.3.2.1 On-Site Support

The contractor shall provide at least one full-time on-site FORTRAN analyst and at least one full-time on-site software systems analyst. These two analysts will be available for consultation from 08:00 to 17:00 local time, Monday through Friday (exclusive of Government holidays).

3.3.2.2 On-Call Support

The contractor shall provide a hardware customer engineer on-call twenty-four (24) hours per day. The offeror also shall provide a software systems analyst on-call during all the hours that the on-site software systems analyst is not on duty at NAS. Both these analysts

will be available for service on-site within two (2) hours when requested.

3.3.3 Training

The contractor shall provide two hundred (200) instructor hours of on-site training of NAS personnel, as required.

3.3.4 Documentation Support

The offeror shall provide to the Government complete documentation of the hardware, the operating system, languages and libraries. On-line documentation shall be provided for frequently used system commands and utilities, and shall be available to users. The Government shall have the right to make copies of documentation for NPSN users, including those off-site.

4 NHT-1 PERFORMANCE SPECIFICATIONS

4.1 Floating Point Computation Benchmark Performance

The Half-Size Testbed shall achieve a performance rate exceeding 1.5 billion 64 bit floating-point operations per second on a suite of benchmarks specified by the Government. The Full-Size Testbed shall achieve a performance rate exceeding 3 billion 64 bit floating-point operations per second on a suite of benchmarks specified by the Government. These performance rates are approximations. The actual test requirements are specified in terms of the wall clock times between initiation and completion of execution. Details of these requirements are given in Appendix A.

4.2 Mass Storage Benchmark Performance

The mass storage subsystem shall be capable of transferring data to and from computational node memory at a rate sufficient to pass the benchmark tests that are required in Sections 4.2.1 and 4.2.2.

4.2.1 This test shall be provided by the government and is described in detail in Appendix A, Section A.5.3. Note that this test requires certain network benchmark tests to be run simultaneously with the mass storage data transfers. These are all components of this mass storage benchmark test.

4.2.2 This test shall consist of a benchmark FORTRAN program to be supplied by the offeror. The detailed requirements for this test are described in Appendix A, Section A.5.4.

4.3 System Availability

4.3.1 Definitions for Availability Specifications

In this section, "available," "operational," "operative" or "up" means that the system can correctly execute the Government's workload. "Inoperative" or "down" refers to a system that cannot correctly execute the Government's workload. The system shall be considered "down" if it at any time fails to complete any of the Government's specified installation tests at the required performance. "Downtime" begins when the Government notifies the vendor's point of contact and reports that the system is down.

4.3.1.1 During the Acceptance Test, "uptime" begins when the system begins executing the Government's workload. The contractor must return the system to an operational status according to a procedure approved by the Government.

4.3.1.2 After completion of the Acceptance Test, "uptime" begins when the vendor officially notifies the Government that the system is ready to return to operational status.

4.3.2 Acceptance Test Requirements

During the Acceptance Test, the system may be allocated to the contractor up to four (4) hours per twenty-four (24) hour day (beginning at midnight) for the purposes of system maintenance. This downtime must be scheduled at least twenty-four (24) hours in advance and must occur between 18:00 and 08:00 local time. The remaining time each day shall be considered scheduled uptime. During the Acceptance Test period, the system must be available to execute the Government's workload at least 80% of scheduled uptime averaged over a 30-day test period.

4.3.3 Production Requirements

After the completion of the Acceptance Test, if the system remains inoperative during scheduled uptime due to a hardware or software malfunction for a total of four (4) hours (whether or not consecutive) during a twenty-four (24) hour day (beginning at midnight), the contractor shall grant a downtime credit to the Government. The downtime credit will accumulate at the rate of 1/20 of a day (5%) for each hour the system is inoperative in excess of four (4) hours. The credit will be prorated for fractional hours. The system will not be regarded as operative until it has been up for two (2) consecutive hours. At that time, the previous two (2) hours will be retroactively counted as uptime.

5 FACILITY PLANNING AND SITE PREPARATION

The contractor shall provide a plan for facility planning and site preparation. The contractor shall be responsible for providing a site preparation plan for the proper installation and operation of the equipment, such as connections from building power sources, including vendor provided motor generator sets to the computer room equipment, etc.

6 MAINTENANCE

6.1 Responsibilities of the Contractor

The contractor shall provide maintenance for the equipment delivered and installed and shall keep the equipment in good operating condition. The contractor shall arrange for maintenance services for all delivered hardware regardless of the manufacturer.

6.2 Preventive Maintenance for Leased Equipment

The Contractor shall specify in writing the frequency and duration of the preventive maintenance required. If a mutually agreed upon schedule for preventive maintenance cannot be established, the Government reserves the right to specify the schedule for performance of preventive maintenance.

6.3 Preventive Maintenance for Purchased Equipment

The Contractor shall specify in writing the frequency, duration, and quality of preventive maintenance. The quality shall be comparable to that provided by the Contractor for identical leased equipment. If a mutually agreed upon schedule for preventive maintenance cannot be established, the Government reserves the right to specify the schedules for performance of preventive maintenance.

7 FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATIONS (FIPS-PUBS)

In accordance with the Federal Information Resources Management Regulations (FIRMR), the following Federal Information Processing Standards Publications (FIPS-PUBS) shall constitute a part of this specification and are applicable to the extent specified herein:

FIPS 1-2	Code for Information Interchange, Its Representations, Subsets and Extensions
FIPS 46	Data Encryption Standard (DES)
FIPS 50	Recorded Magnetic Tape for Information Interchange, 6250 CPI (246 cpmm), Group Coded Recording
FIPS 60-2	I/O Channel Interface
FIPS 61-1	Channel Level Power Control Interface
FIPS 62	Operational Specifications for Magnetic Tape Subsystems
FIPS 69	Federal Standard FORTRAN
FIPS 79	Magnetic Tape Labels and File Structure for Information Exchange
FIPS 81	Data Encryption Standard (DES) Modes of Operation
FIPS 86	Additional Controls for use with American National Standard Code for Information Interchange
FIPS 112	Accepted Practices for Password Usage
FIPS 113	Data Authentication (DES)
FIPS 130	Intelligent Peripheral Interface (IPI)
FIPS 149	Government Open Systems Interconnection Profile (GOSIP)
FIPS 151-1	POSIX: Portable Operating System Interface for Computer Environments (IEEE 1003.1-1988)

8 REFERENCES

8.1 System V Interface Definition Third Edition, Volume I, II, III, IV, Order #320.135, American Telephone and Telegraph, Inc., AT&T Customer Information Center, 2833 North Franklin Road, Indianapolis, IN 46219.

8.2 Unix Time-Sharing System, Unix Programmer's Manual, 4.3 Berkeley Distribution, Volume 1, November 1986, University of California, Berkeley, CA 94708.

8.3 The Internet Protocol Transition Handbook, 1982 Network Information Center, SRI International, Menlo Park, CA 94025.8.4. American National Standard Programming Language Fortran, ANSI X3.9 - 1978, American National Standards Institute, 1430 Broadway, New York, NY 10018.

8.4 American National Standard Programming Language FORTRAN, ANSI X3.9 - 1978, American National Standards Institute, 1430 Broadway, New York, NY, 10018.

8.5 Network Queueing System Software. Submission Number ARC 11750. Available from Cosmic Computer Services Annex, University of Georgia, Athens, GA 30602, (404) 542-3265.

8.6 Network Working Group Request for Comment 950, "Internet Standard Subnetting Procedures," (by Mogul, J. and Postel, J.). USC/Information Sciences Institute, Los Angeles, CA, August 1985.

8.7 American National Standard Programming Language C, Document #X3.159-1989, American National Standards Institute, 1430 Broadway, New York, NY, 10018.

8.8 Ultra Network Technologies, 101 Daggett Drive, San Jose, CA, 95134, USA

8.9 Requirements for Internet Hosts -- Communication Layers, RFC 1122, October 1989, Internet Engineering Task Force. Network Information Center, SRI International, Menlo Park, CA, 94025.

8.10 Requirements for Internet Hosts -- Application and Support, RFC 1123, October 1989, Internet Engineering Task Force. Network Information Center, SRI International, Menlo Park, CA, 94025.

8.11 "External Data Representation Protocol Specification" in Networking on the Sun Workstation, part #800-1324-03, Sun Microsystems, Inc., 2550 Garcia Avenue, Mountain View, CA 94043.

8.12 "Remote Procedure Call Protocol Specification" in Networking on the Sun Workstation, part #800-1324-03, Sun Microsystems, Inc., 2550 Garcia Avenue, Mountain View, CA 94043.

8.13 "ONC/NFS Protocol Specifications and Services Manual", Part No. 800-3084-10, Rev A, of 26 August 1988. Sun Microsystems, Inc., 2550 Garcia Avenue, Mountain View, CA 94043.

8.14 X Windows Version 11, Release 4
X Window System Protocol, X Version 11, Release 4
Xlib - C Language X Interface, X Version 11, Release 4
X Toolkit Intrinsics C Language Interface, X Version 11, Release 4
Bitmap Distribution Format, Version 2.1
Inter-Client Communication Conventions Manual, Version 1.0
Compound Text Encoding, Version 1.1
X Logical Font Description Conventions, Version 1.3
X Display Manager Control Protocol, Version 1.0
X11 Nonrectangular Window Shape Extension, Version 1.0
Specifications available from:
MIT X Consortium
Laboratory for Computer Science
545 Technology Square
Cambridge, MA 02139

8.15 High Performance Parallel Interface (HiPPI)
Draft proposed HiPPI-PH ANSI Standard X3T9.3/88-023
Draft proposed HiPPI-DF ANSI Standard X3T9.3/89-013
American National Standards Institute, 1430 Broadway, New York, NY, 10018.

8.16 IEEE Standard for Binary Floating Point Numbers, ANSI/IEEE Standard 754-1985, IEEE, New York, 1985.

8.17 PCF Fortran Extensions -- Draft Document, Revision 2.11, March 18, 1990; PCF, c/o Kuck and Associates, 1906 Fox Drive, Champaign, Illinois 61820.

8.18 D. Bailey, J Barton, T. Lasinski, and H. Simon, eds. *The NAS Parallel Benchmarks, Revision 2.* Technical Report RNR-91-002, NASA Ames Research Center, Moffett Field, CA 94035-1000, July 2, 1991.

8.19 *UNIX System V Release 4 Programmers Guide: ANSI C and Programming Support Tools.* 1990, Prentice-Hall, Inc. New Jersey.

APPENDIX A
DETAILS OF THE BENCHMARK TESTS

A.1 INTRODUCTION

A.1.1 General Information

Appendix A establishes the specific procedures and performance requirements for the execution of the NHT-1 benchmark tests. The "Benchmark Package" (a tape containing the source code, input data, and sample output files for single processor, scaled-down implementations the benchmarks) will be provided to potential offerors who specifically request the benchmark package in writing to Ames Research Center.

The benchmark tests for Testbed proposals are described below. These tests, collectively identified as the "RFP Response Test," are to be performed by the offeror in response to the RFP, and the results of these tests are to be included in the offeror's proposal. If the offeror's proposal is selected for contract award, then the exact same set of benchmark tests is to be repeated at the vendor's site prior to shipment. This set of tests is denoted the "Pre-Shipment Test." The vendor is required to demonstrate prior to shipment that the proposed system meets or exceeds the performance claimed in the response to the RFP.

All questions concerning the running of the benchmark programs should be directed to the Contracting Officer. Every effort will be made to provide answers to questions in a timely manner. Questions and answers will be disseminated to all offerors. Ames Research Center will not reimburse offerors for machine time, programming efforts, or any other expenses incurred in preparing and performing these benchmark tests.

The benchmark programs and data are the property of NASA and are not to be used for any purpose other than the specified demonstrations. No copies of these programs or data may be disseminated in any form to outside parties, and all copies of the programs and accompanying documentation shall be returned to Ames Research Center upon notice of non-selection. The Contracting Officer shall appoint a Contracting Officer's Technical Representative (COTR) to make final decisions regarding evaluation of the results of the benchmark programs and the demonstrations. NASA may require the offeror to demonstrate any proposed hardware or software specifically described in their proposal, for the purpose of verifying that the system performs as proposed. The offeror will be given adequate notice of any such additional demonstration requirements. No reimbursement will

be made for these additional demonstrations. Offerors may request, and the Government reserves the right, to delay the demonstration of the benchmark tests, if it is judged to be in the best interest of the Government.

A.1.2 Testbed Type and the Benchmark Tests

The benchmark tests presented below apply to proposals for both the Half-Size and Full-Size Testbeds. Where different performance levels are required depending on the type of Testbed, these differences are explicitly noted. Unless otherwise explicitly stated in the appropriate section, the requirements for each type of Testbed are identical.

A.2 INSTRUCTIONS FOR THE RFP RESPONSE TEST

Passage of the RFP Response Test requires that the offeror demonstrate successful completion of the benchmark tests specified in Sections A.4, A.5, A.5.1-4, A.6.1-3. The results of the offeror's RFP Response Test are to be included in the offeror's response. The RFP Response Test results must be verified by the Government as specified in Section A.4 prior to selection of successful proposals.

A.3 INSTRUCTIONS FOR THE PRE-SHIPMENT TEST

Passage of the Pre-Shipment Test requires that the offeror demonstrate prior-to-shipment performance on the benchmark requirements specified in Sections A.4, A.5, A.5.1-4, A.6.1-3 that meets or exceeds the level proposed for the offeror's RFP Response. The Pre-Shipment Test must be performed on the actual hardware and software configuration intended for delivery to NAS. The Pre-Shipment Test results must be verified by the Government as specified in Section A.4 prior to shipment.

A.4 GENERAL INSTRUCTIONS FOR RUNNING THE BENCHMARK TESTS

The time, location, and participants in the Benchmark Test demonstrations must be mutually agreed to by both NASA and the offeror. The offeror shall provide a suitable private conference room for use by the Ames Research Center participants. The participants provided by the offeror shall include persons knowledgeable in both the hardware and software systems to answer questions that arise in the course of the demonstration. For all the tests described in this section that require execution of source programs other than standard system software, the compilation and linking of this code must be

performed, as a prelude to the test, in the presence of NASA benchmark representatives, on the service node subsystem of the NHT-1. The same version of the compiler and linker must be used for all tests. Once the test is ready to execute, it shall be started only after a signal is given by the officially designated NASA benchmark representative.

Where the performance requirement is stated in terms of "wall clock time" this time is defined as the elapsed continuous time-of-day from the beginning of the test execution until the completion of the test. No other jobs may be executing in the system when a demonstration test is being performed, and no system tuning or other human intervention may occur during the execution of the test. In the case of an interruption of a test run due to an operating system failure, NASA representatives, after consultation with offeror's participants, will decide whether or not to rerun the test. In the case of an interruption due to circumstances totally beyond the offeror's control (such as a power outage), a rerun will be allowed. Any rerun shall at the COTR's option be performed from the beginning of the test.

After the completion of any of these test demonstrations, the offeror must, upon request, provide to NASA representatives the following:

1. A complete description of all system software used, including version numbers.

2. A paper source listing of the actual programs demonstrated in the test.

3. A paper listing of the output files for the test.

4. A separate paper listing of the differences between the output files of the test and the reference output files, if any, provided by NASA.

5. A disk file copy of items 1, 2, 3 and 4.

The offeror is also strongly encouraged to provide a description of not more than one page length detailing with references the algorithm used, and a description of not more than one page detailing the data layout scheme used for each benchmark.

Verification of the Benchmark Test results by the Government will be made based on a technical analysis of the Benchmark Test demonstrations.

A.5 OPERATING SYSTEM VALIDATION, DATA TRANSFER AND NETWORKING TESTS

There are four separate tests involved in passing Appendix A Section A.5. The tests are the System V Verification Suite (SVVS) (Section A.5.1), the Functional Performance Test (Section A.5.2), the

Throughput Performance Test (Section A.5.3) and the FORTRAN Mass Storage File Tests (Section A.5.4). Each test must be run without modification and successfully completed. The operating system configuration under which these tests are run must be identical to the configuration under which all other Benchmark Tests are run.

A.5.1 System V Verification Suite (SVVS)

The System V Interface Definition (SVID) [8.1] must be run on the service node operating system and dedicated network service nodes. This functionality will be verified by successful completion of the System V Verification Suite. Compatibility failures must be noted, and any exceptions allowed must be approved by the Government.

A.5.2 Functional Performance Test

The Functional Performance Test is included in the Benchmark Package in the form of a shell archive. The offeror shall be required to use the Bourne shell to unpack the archive, use the install shell script to set up the test, and then use the run shell script to run the test to completion. The install shell script asks the vendor to supply the pathname to the networking I/O library and the hostnames to be used in the ethernet HiPPI/Ultra functional tests. It also asks for the number of files to include in the run test, as well as the pathname of each file. The script finally asks for configuration information needed to correctly compile the functional test. It then attempts to install the components for the functional performance test and the I/O throughput test. The install script reports success or failure. Once the install script has completed, the functional test shall be run. This test verifies that a Berkeley Standard Distribution (BSD) socket library is available and functional, and that it is possible to transfer files using the ftp command at the network addresses specified. In addition to ftp, a Government implementation of the echo protocol will be used to verify BSD functionality. The run scripts reports success or failure. This test must be run on all nodes that provide network access.

A.5.3 Throughput Performance Test

This test is included in the Bourne shell archive referenced in Section A.5.2 above. This test uses the functionality demonstrated by the test in Section A.5.2. It also has an install and run shell script. The install shell script asks for the hostnames to use as well as pathnames to use for source and destination files. This script then creates a database file used in the actual throughput performance test. It is likely that the target hardware configuration provided will not support this test. This test

may be modified with the permission of the COTR in order to run on the hardware provided. In addition to the network test described below and provided by the Government in the Benchmark Package, the test required in Section 4.2.2 will be incorporated into this test by the vendor so as to run at the same time as the networking portion of this benchmark test. All modifications must be approved by the COTR.

The run shell script builds all of the source files. When this has completed, the current system time is printed out indicating the start of the benchmark portion of the test. The required file transfers are then started at approximately the same time by creating a background process for each transfer. The vendor's FORTRAN I/O benchmark required in Section 4.2.2 shall begin the "READ TEST" portion of the test at this point followed by the "WRITE TEST" portion of the test. When all of the file transfers have completed, including the test from Section 4.2.2, the system time is once again printed out. All of these file transfers must be successfully completed without error within a single 120 second wall clock time interval. This interval is measured by the difference in wall clock time between the first printing of the system time and the final printing of the system time. A log file for each file transfer is also created by the run shell script. These logs will be examined to help determine the success of the benchmark. File transfer through the network shall be done with the vendor's *ftp* program.

The required benchmark rates are given below in Table 1 for each type of file transfer. Each individual rate must be met or exceeded to pass the above benchmark wallclock time limit.

Table 1

Transfer Type	Number of Streams	Source File Size per Stream (billion bytes)	Minimum Required Transfer Rate per Stream (million bytes per second)
HiPPI/Ultra	1	4.8	40.0
Ethernet	2	0.30	0.25

The run shell script shall be executed twice. During the first execution, the ethernet and HiPPI/Ultra network interfaces shall be disconnected from the system to demonstrate that the connection is through the wire loop back. This run shall generate vendor-dependent diagnostics in the log. If it appears to succeed, or if the diagnostics do not appear, then the run is considered a failure. Immediately afterwards, the networks shall be reconnected, without modifying the running system in any way, and the run shell script shall be executed again. This time

it shall complete successfully within the specified time. The network portion of the run shell script consists of two executions of the ftp program, each using a different pair of ethernet adapters, one as output and one as input. Simultaneously, one execution of the ftp program using one pair of HiPPI/Ultra interfaces shall occur. Throughput is measured to disk in all three cases.

A.5.4 FORTRAN Mass Storage File Test

The offeror shall prepare a FORTRAN program that shall first create mass storage files totaling at least 10 billion bytes. The program shall then perform the "READ TEST" portion of the test which is to read into computational node memory all 10 billion bytes (overwriting the data in memory). The program shall then perform the "WRITE TEST" portion of the test which is to write out at least 10 billion bytes of data, overwriting the original 10 billion bytes of data. For this test, the "READ TEST" portion of the test shall complete in 60 seconds wall clock time or less. The "WRITE TEST" portion of the test shall complete in 60 seconds wall clock time or less. The entire test shall complete in 120 seconds wall clock time or less.

A.6 MULTIPLE PROCESSOR PERFORMANCE TEST

A.6.1 Introduction

The Multiple Processor Performance Test is based on the "NAS Parallel Benchmarks" [8.18]. The performance level required of a Testbed for this test differs depending on whether the proposal is for a Half-Size or Full-Size Testbed.

A.6.2 Language Rules

The Multiple Processor Performance Test must be programmed using the Fortran-77 language, with a number of approved extensions. The complete language rules are given in [8.18], Section 1. All of these rules apply to this test, with the exception that coding the problems in the C programming language is not allowed here.

A.6.3 Multiple Processor Performance Test Details

The Multiple Processor Performance Test will consist of the problems defined in the NAS Parallel Benchmarks [8.18], with the problem sizes increased to the sizes listed in Table 2 below. In Table 2, the column headed "Size" gives the values of the key problem parameters. The

column headed "Memory" gives approximate memory requirements in millions of 64-bit words, based on one processor computer implementations. The column headed "FP Ops" gives approximate counts of floating point operations, based on one processor computer implementations. The sample code files included in the Benchmark Package contain instructions on how the parameters in these one processor programs may be increased to the problem sizes listed in Table 2.

Table 2

NHT-1 Multiple Processor Performance Pre-Shipment Test

Benchmark Description	Size	Memory	FP Ops
1 Embarrassingly Parallel	2^{30}	1	8.9×10^{10}
2 Simplified Multigrid	$512^2 \times 256$	480	3.2×10^{10}
3 Conjugate Gradient	7.4×10^6	50	6.0×10^9
4 3-D FFT PDE Solver	512×256^2	500	3.2×10^{10}
5 Integer Sort	2^{25}	128	1.0×10^{10}
6 LU Solver	102^3	26	4.5×10^{11}
7 Pentadiagonal Solver	102^3	22	4.9×10^{11}
8 Block Tridiagonal Solver	102^3	22	7.5×10^{11}

The results of this test are subject to review and verification by the COTR. Before any of these kernel tests are performed, the source code produced by the vendor for all of the benchmark programs must be compiled and linked, using a single version of the Fortran compiler and linker for the computational nodes. Compiling and linking the entire set of Multiple Processor Performance Test source files must be completed within 30 minutes wall clock time. The individual computational tests described below must be performed with the executables files resulting from this compile-link operation. The operating system must remain up during compilation, linking, and execution of all the benchmark codes.

The performance p_i of the proposed system on each of the eight kernel benchmarks will be computed as follows:

$$p_i = \frac{F_i}{t_i} \qquad i = 1, 2, 3, \ldots, 8$$

where F_i is the number of floating point operations (FP Ops) for the ith kernel listed in Table 2, and t_i is the measured elapsed runtime in seconds of the i th kernel.

The minimum performance requirement for the Multiple Processor Performance Test is that for each of the eight kernels the performance

p_i must exceed 1.5 billion FP Ops per second for the Half-Size Testbed, or 3.0 billion FP Ops per second for the Full-Sized Testbed.

Proposals that meet this minimum performance requirement will then be evaluated by the rank of their overall multiple processor figure of merit P. P is computed as follows:

$$P = \frac{2}{25}*(\frac{t1}{T1} + \frac{t2}{T2} + \frac{t3}{T3} + \frac{t4}{T4} + \frac{t5}{T5}) + \frac{1}{5}*(\frac{t6}{T6} + \frac{t7}{T7} + \frac{t8}{T8})$$

where $t_1, t_2, ..., t_8$ are defined above, and $T_1, T_2, ..., T_8$ are computed in the following manner.

The proposals will be divided at the time of evaluation into two groups: one for proposals for Half-Size Testbeds and the other for proposals for Full-Size Testbeds. For each benchmark kernel and group, the corresponding T_i is defined to be the minimum elapsed runtime measured in seconds for that kernel in any of the verified vendor proposals in the group. Thus proposals for Half-Size Testbeds will have an overall multiple processor figure of merit P computed on the basis of the group of Half-Size Testbed proposals, and likewise the proposals for Full-Size Testbeds will have P computed on the basis of the group of Full-Size Testbed proposals.